THE KALMAN FILTER: OPTIMAL STATE ESTIMATION IN THE PRESENCE OF NOISE — lectures 1 and 2

John M. Finn

9th February 2004

UQWG seminar, Feb 5, 2004

Outline

- Counting statistics with equal σ_i by least squares approach. Minimum variance. Recursive nature.

- Counting statistics with unequal σ_i. Least squares, minimum variance approach. Recursive nature.

- Linear process with measurement noise only – estimating initial vs. current state.

- Random walk with zero measurement noise. Estimating the initial position.

- Random walk, estimating the current position.

- Random walk with measurement noise, estimating the current state.

- Preview of next lecture.

Counting statistics; sample mean and variance – equal σ_i^2

$$x_i = x_0 + \xi_i \quad y_i = x_i,$$

$< \xi_i >= 0$, $< \xi_i \xi_j >= \sigma_i^2 \delta_{ij} = \delta_{ij}$ and gaussian distribution by Bayes' theorem

$$f(x_0(n)|y_1, ..., y_n) \propto f(y_1, ..., y_n|x_0(n)) \times \frac{prior}{normalization}$$

$$\sim \Pi_{i=1}^{n} e^{-\frac{1}{2\sigma_i^2}[x_i - x_0(n)]^2} = e^{-\sum_{i=1}^{n} \frac{1}{2\sigma_i^2}[x_i - x_0(n)]^2}$$

Maximum likelihood

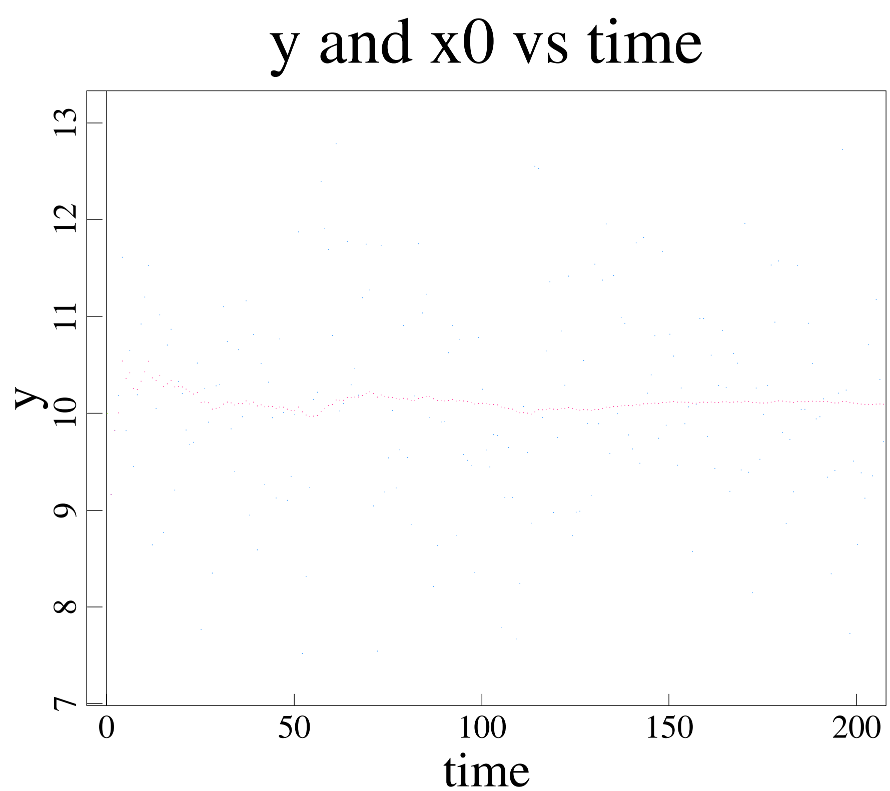
$$\chi^2(n) = -\ln f \; = \; \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2}.$$

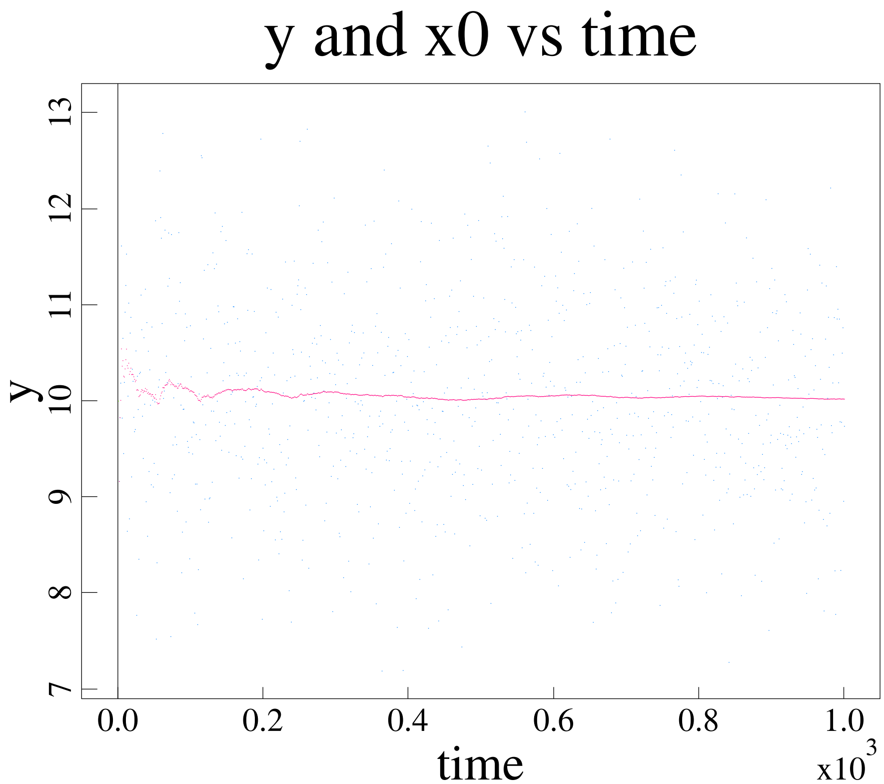
$\partial\chi^2/x_0(n) = 0 \Rightarrow$ state estimate

$$x_0(n) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

The variance *of the estimate* at this stage is (uncorrelated)

$$V(n) = \sigma^2(n) = \sum_{i=1}^{n} \sigma_i^2 (\partial x_0(n)/\partial x_i)^2 = \frac{1}{n},$$




y and x0 vs time
y
13
12
11
10
9
8
7
0.0
0.2
0.4
0.6
0.8
1.0
time
x10^3

Minimum variance approach

$$x_0(n) = \sum_{i=1}^{n} \rho_i x_i,$$

with $\sum_{i=1}^{n} \rho_i = 1, \ \ \sigma^2 = 1$

$$V(n) = \sum_{i=1}^{n} \rho_i^2 \qquad V^*(n) = \sum_{i=1}^{n} \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i$$

$\partial V(n)/\partial \rho_k = 0 \ \ \Rightarrow$

$$\rho_k = \frac{\lambda}{2},$$

or $\rho_k = 1/n$ for all k – $x_0(n) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad V(n) = 1/n$

Recursive Kalman filter form

$$(n+1)x_0(n+1) = \sum_{i=1}^{n} x_i + x_{n+1}$$

$$x_0(n+1) = \frac{n}{n+1}x_0(n) + \frac{1}{n+1}x_{n+1}$$

or

$$x_0(n+1) = x_0(n) + K_n[x_{n+1} - x_0(n)],$$

Kalman gain K_n

$$K_n = \frac{1}{n+1}.$$

$$\frac{1}{K_n} = \frac{1}{K_{n-1}} + 1 \quad or \quad K_n = \frac{K_{n-1}}{1 + K_{n-1}}$$

$$K_n = V(n+1),$$

Counting statistics for unequal σ_i^2

Uncorrelated but different confidence: $<\xi_i \xi_j>= \sigma_i^2 \delta_{ij}$

$$\chi^2 = \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2\sigma_i^2}$$

$\partial W / \partial x_0(n) = 0 \quad \Rightarrow$

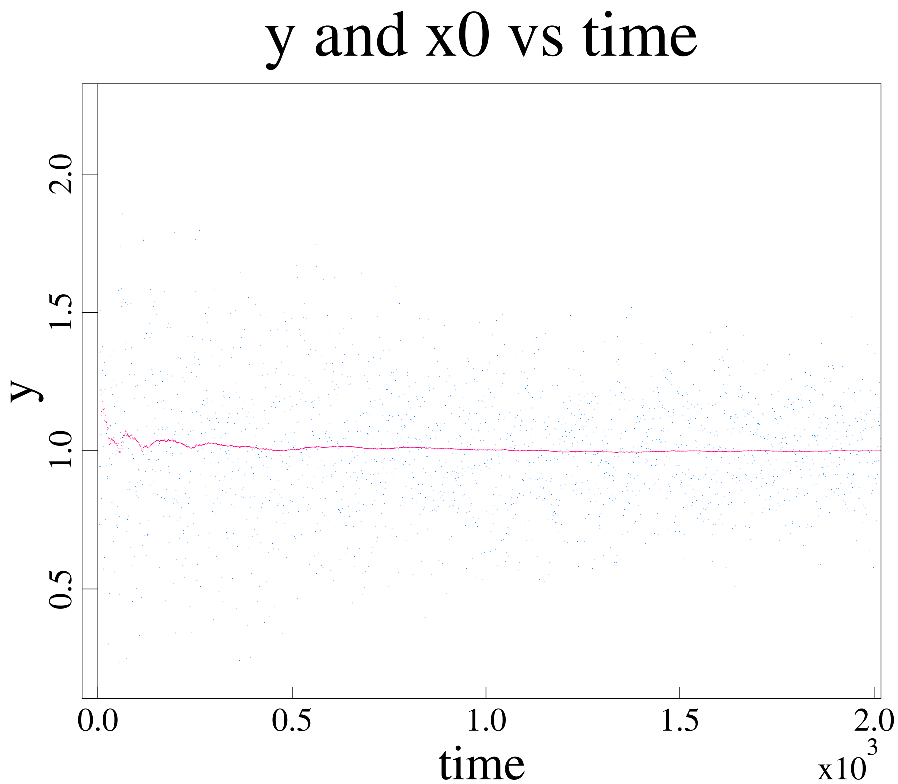
$$x_0(n) = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}.$$

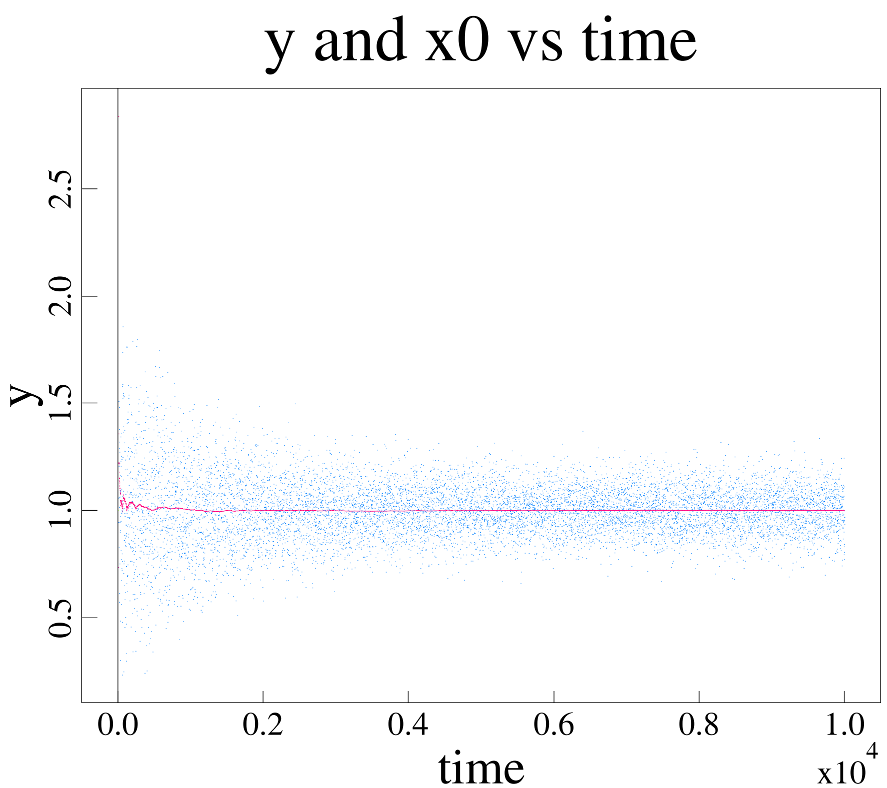
Example: $(x_1, x_2, x_3), x_4$

Take $z_1 = (x_1 + x_2 + x_3)/3,\ z_2 = x_4$. $\sigma_1^2 = 1/3,\ \sigma_2^2 = 1$

Then $x_0(4) = \left(z_1/\sigma_1^2 + z_2/\sigma_2^2\right) / \left(1/\sigma_1^2 + 1/\sigma_2^2\right)$

$$= (x_1 + x_2 + x_3 + x_4)/4$$


y and x0 vs time
y
2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
time
x10^3


y and x0 vs time
y
2.5
2.0
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
time
x10^4

$$V(n) = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2},$$

Again, take

$$x_0(n) = \sum_{i=1}^{n} \rho_i x_i,$$

with $\sum_{i=1}^{n} \rho_i = 1$

$$V^*(n) = \sum_{i=1}^{n} \sigma_i^2 \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i$$

$$\rho_k = \frac{\lambda}{2\sigma_k^2} = \frac{1/\sigma_k^2}{\sum_i 1/\sigma_i^2}.$$

Same.

Recursive Kalman filter form for unequal σ_i^2

$$x_0(n+1) = x_0(n) + K_n[x_{n+1} - x_0(n)]$$

with

$$K_n = \frac{1}{\sigma_{n+1}^2 \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} + \frac{1}{\sigma_{n+1}^2} \right)} = \frac{1}{\sigma_{n+1}^2 \sum_{i=1}^n \frac{1}{\sigma_i^2} + 1}.$$

$K_n = V(n+1)/\sigma_{n+1}^2$, and

$$\frac{1}{K_n} = \left(\frac{\sigma_{n+1}^2}{\sigma_n^2} \right) \frac{1}{K_{n-1}} + 1 \quad \text{or} \quad K_n = \frac{K_{n-1}}{K_{n-1} + \sigma_{n+1}^2/\sigma_n^2},$$

The recursion in terms of the variance

$$\frac{1}{V(n+1)} = \frac{1}{V(n)} + \frac{1}{\sigma_{n+1}^2}.$$

with $K_n = V(n+1)/\sigma_{n+1}^2$ K_n tends to decrease with n (more data)

If $\sigma_{n+1}^2 < \sigma_n^2$, then K_n will be larger than if $\sigma_{n+1}^2 > \sigma_n^2$

One dimensional example of estimating the initial state and the current state

Simple stochastic system with measurement noise

$$x_{k+1} = \gamma x_k,$$

$$y_k = x_k + \eta_k.$$

$< \eta_k \eta_l >= \delta_{kl}$

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n} (\gamma^k x_0 - y_k)^2,$$

$\partial \chi^2 / \partial x_0 = 0$ gives

$$x_0(n) = \frac{\sum_{k=1}^{n} \gamma^k y_k}{\sum_{k=1}^{n} \gamma^{2k}}.$$

$\gamma > 1$...weighted toward recent results, $\gamma < 1$... weighted toward initial results. Recursive form

$$x_0(n+1) = x_0(n) + \frac{\gamma^{n+1}}{\sum_{k=1}^{n} \gamma^{2k} + \gamma^{2n+2}} \left(y_{n+1} - \gamma^{n+1} x_0(n) \right).$$

An estimate of x_n rather than x_0.

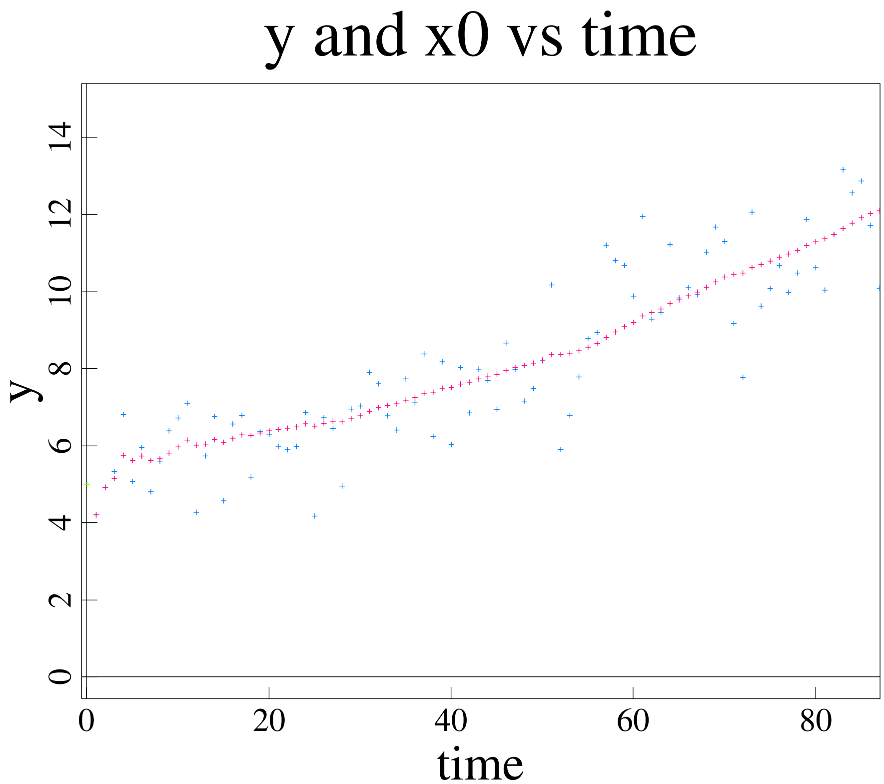
$$\chi^2 = \frac{1}{2}\sum_{k=1}^{n}(\gamma^{k-n}x_n - y_k)^2,$$

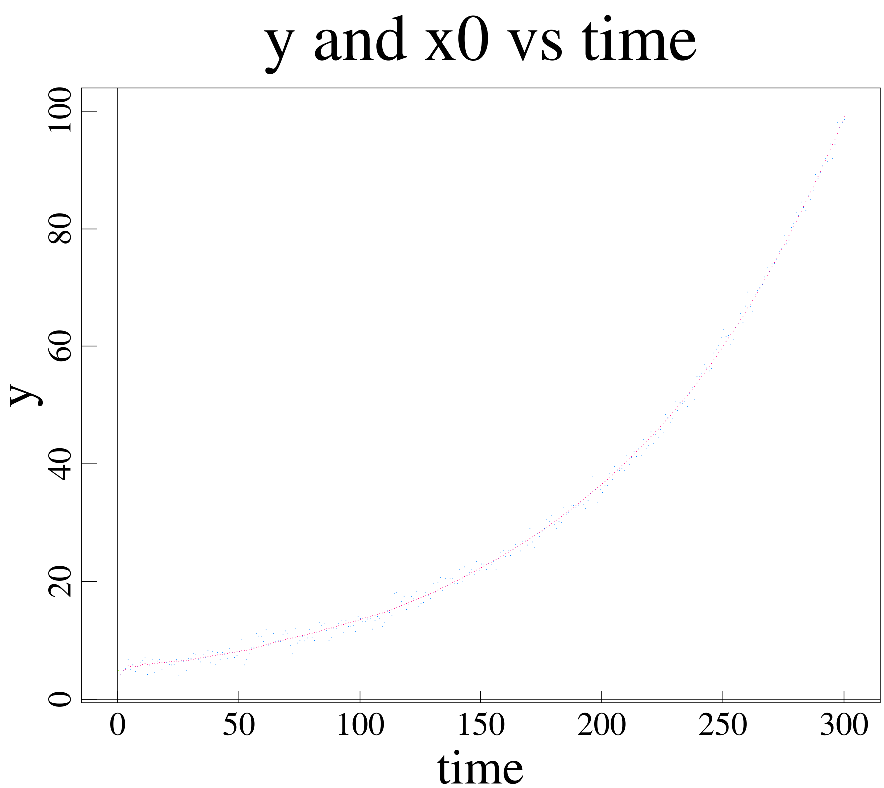
$$x_n(n) = \frac{\sum_{k=1}^{n}\gamma^{k-n}y_k}{\sum_{k=1}^{n}\gamma^{2k-2n}} = \gamma^n x_0(n),$$

Exactly what you might guess. Recursive form:

$$x_{n+1}(n+1) = \gamma x_n(n) + \frac{1}{\sum_{k=1}^{n}\gamma^{2k-2n-2} + 1}\left(y_{n+1} - \gamma x_n(n)\right).$$

Notice $K_n^{x_n(n)} = \gamma^{n+1}K_n^{x_0(n)}$.


y and x0 vs time
y
0
2
4
6
8
10
12
14
0
20
40
60
80
time


y and x0 vs time
y
0
20
40
60
80
100
0
50
100
150
200
250
300
time

Random walk with zero measurement error – estimating the *initial* position

Counting statistics, with only measurement noise, is:

$$x_{k+1} = x_k,$$

$$y_k = x_k + \eta_k$$

Random walk problem (Wiener process, Brownian motion), with only dynamical noise:

$$x_{k+1} = x_k + \xi_k,$$

$$y_k = x_k.$$

$< \xi_k >= 0$, $< \xi_k \xi_k >= \sigma_0^2 \delta_{kl}$. To estimate the initial position. Ship wrecks at x_0 – to find the ship.

$$y_k = x_0 + \sum_{i=0}^{k-1} \xi_i = x_0 + \zeta_k.$$

ζ_k has $< \zeta_k >= 0$

And $n \times n$ covariance matrix

$$C_{kl} =< \zeta_k \zeta_l >= C_{kl} =< \zeta_k \zeta_l >$$
$$= \sum_{i=0}^{k} \sum_{j=0}^{l} < \xi_i \xi_j > = \sigma_0^2 \min(k, l),$$

i.e.

$$\mathsf{C} = \sigma_0^2 \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & 3 & \cdots & 3 \\ 1 & 2 & 3 & 4 & \cdots & 4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & 4 & \cdots & n \end{bmatrix}.$$

Least squares in terms of the inverse of the covariance matrix $\mathsf{D} = \mathsf{C}^{-1}$

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} \zeta_k D_{kl} \zeta_l = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} (y_k - x_0) D_{kl} (y_l - x_0).$$

$\partial\chi^2/\partial x_0 = 0 \quad \Rightarrow$

$$x_0(n) = \frac{\sum_{k=1}^{n}\sum_{l=1}^{n} D_{kl} y_l}{\sum_{k=1}^{n}\sum_{l=1}^{n} D_{kl}},$$

$$\mathsf{D} = \sigma_0^{-2} \begin{bmatrix} 2 & -1 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & \dots \\ 0 & -1 & 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

In the estimate, the value of σ_0^2 cancels.

$$\sum_{kl} D_{kl} = 1, \sum_{kl} D_{kl} y_l = y_1 \quad x_0(n) = y_1$$

$$V(n) = \sum_{ij} \frac{\partial x_0(n)}{\partial y_i} C_{ij} \frac{\partial x_0(n)}{\partial y_j} = C_{11} = 1.$$

Recursive $x_0(n+1) = x_0(n) + K_n[y_{n+1} - x_0(n)]$ with $K_n = 0$.

Try minimum variance again

$$x_0(n) = \sum_{i=1}^{n} \rho_i y_i$$

$$V^*(n) = \sum_{ij} C_{ij}\rho_i\rho_j - \lambda \sum_i \rho_i;$$

$\partial V(n)/\partial \rho_k = 0 \quad \Rightarrow$

$$\sum_j C_{kj}\rho_j = \lambda/2$$

$$\rho_i = \frac{\lambda}{2}\sum_j D_{ij} \;\; or \;\; \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{pmatrix} = \frac{\lambda}{2}\mathsf{D}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

$$\frac{\lambda}{2} = \frac{1}{\sum_{ij} D_{ij}}; \;\; \rho_i = \frac{\sum_j D_{ij}}{\sum_{ij} D_{ij}}.$$

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$x_0(n) = y_1$$

$V(n) = \sum_{ij} C_{ij}\rho_i\rho_j = C_{11} = 1$. A third approach – next lecture.

Random walk with zero measurement noise – estimating the *current* position

$$x_{k+1} = x_k + \xi_k,$$
$$y_k = x_k.$$

Ship wrecks at x_0, but we wish to find the position of the *survivor*.

$$y_k = x_n - \sum_{i=k}^{n-1} \xi_i = x_n(n) - \zeta_k,$$

$< \xi_i >= 0$, $< \xi_i \xi_j >= \sigma_0^2 \delta_{ij}$ $\qquad \zeta_k^{new} = \zeta_n^{old} - \zeta_k^{old}$.

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} \zeta_k D_{kl} \zeta_l = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (x_n - y_k) D_{kl} (x_n - y_l).$$

$$C_{kl} = < \zeta_k \zeta_l > = \sum_{i=k}^{n-1} \sum_{j=l}^{n-1} < \xi_i \xi_j >$$

$$= \sigma_0^2 \sum_{i=k}^{n-1} \sum_{j=l}^{n-1} \delta_{ij} = \sigma_0^2 [n - \max(k, l)],$$

$$\mathsf{C} = \sigma_0^2 \begin{bmatrix} n-1 & n-2 & n-3 & \cdots & 1 \\ n-2 & n-2 & n-3 & \cdots & 1 \\ n-3 & n-3 & n-3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

$$\mathsf{D} = \sigma_0^{-2} \begin{bmatrix} 1 & -1 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & \dots \\ 0 & -1 & 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix},$$

$$x_n(n) = y_{n-1},$$
$$V(n) = 1.$$

Recursive $x_0(n+1) = x_0(n) + K_n[y_n - x_0(n)]$ with $K_n = 1$ now.

Random walk with measurement noise – estimating the current position

Estimating the current position of the shipwreck survivor

$$x_{k+1} = x_k + \xi_k,$$

$$y_k = x_k + \eta_k.$$

Solve for y_k in terms of x_n:

$$y_k = x_n - \sum_{i=k}^{n-1} \xi_i + \eta_k = x_n - \zeta_k + \eta_k.$$

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n} \sum_{l=1}^{n} (x_n - y_k) D_{kl} (x_n - y_l),$$

D = C^{-1}, with $C_{kl} =< (-\zeta_k + \eta_k)(-\zeta_l + \eta_l) >$. Again using$< \xi_k \xi_l >= \sigma_0^2 \delta_{kl}$, $< \eta_k \eta_l >= \sigma_1^2 \delta_{kl}$, $< \zeta_k \eta_l >= 0$ we have, for $k = 1, ..., m$

$$C_{kl}^{(n)} = \sigma_0^2[n - \max(k, l)] + \sigma_1^2 \delta_{kl},$$

or

$$\mathsf{C}^{(n)} = \sigma_0^2 \begin{bmatrix} n-1 & n-2 & n-3 & \cdots & 0 \\ n-2 & n-2 & n-3 & \cdots & 0 \\ n-3 & n-3 & n-3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$+\sigma_1^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$= \sigma_0^2 \mathsf{C}^{(0,n)} + \sigma_1^2 \mathsf{C}^{(1,n)} = \mathsf{C}^{(n)}.$$

NEXT LECTURE

- Probabilistic (Bayesian) approach
- Application to higher dimension, with dynamical and measurement noise
- Application to control theory
- Application to nonlinear problems – the extended Kalman filter

Outline - Second Lecture

- Review of previous lecture
- Estimation of M correlated variables. Alternate method based on the trace of the covariance matrix.
- Alternate method for the random walk with zero measurement noise. Estimating the initial or current position
- Probabilistic (Bayesian) approach
- Alternate method, for random walk *with* measurement noise added
- Higher dimensional stochastic process with measurement noise
- Application to control theory – the separation theorem
- Nonlinear stochastic systems – the Extended Kalman Filter

REVIEW: ESTIMATING A SCALAR VARIABLE

Measurement of a scalar – measurement noise but no dynamical noise

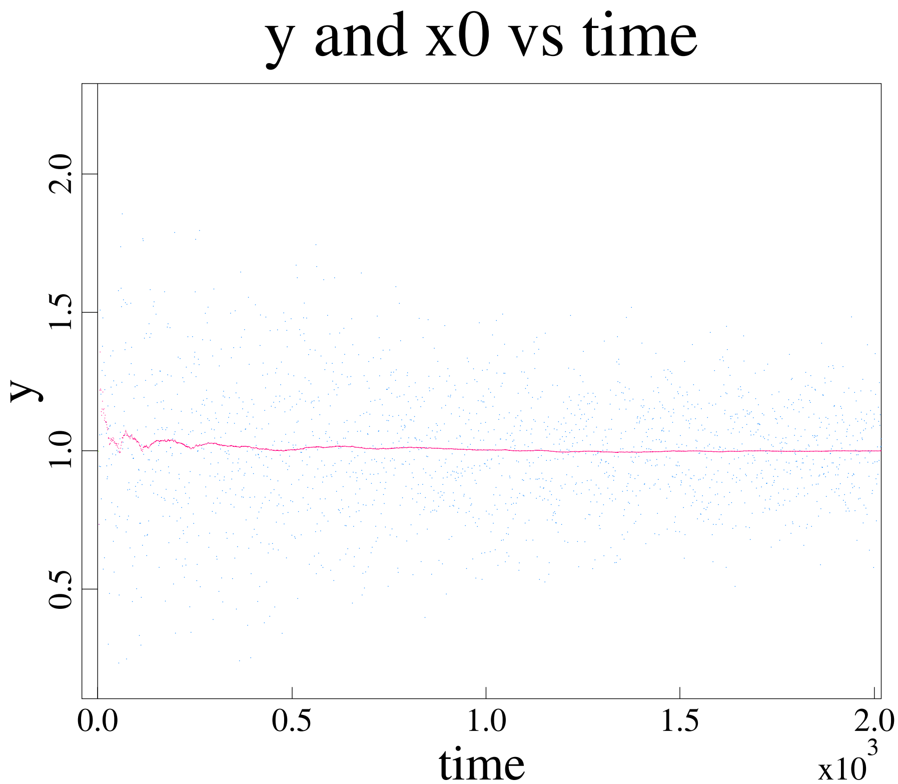
$$\chi^2 = \sum_{i=1}^{n} \frac{[x_i - x_0(n)]^2}{2\sigma_i^2}$$

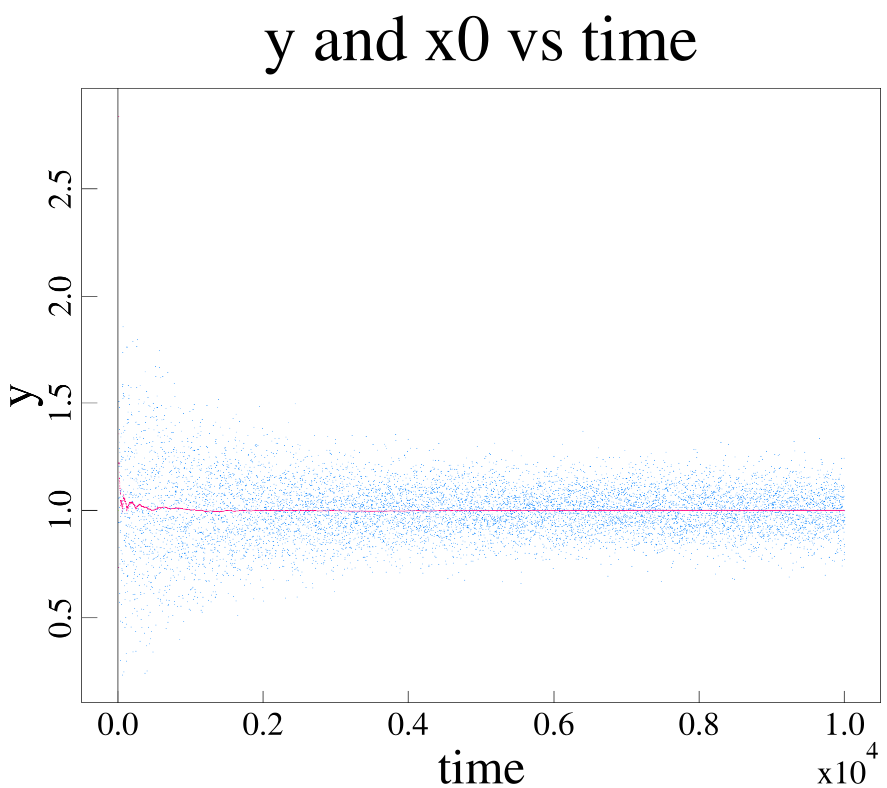
$$x_0(n) = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

Minimum variance form $x_0(n) = \sum_{i=1}^{n} \rho_i x_i$, with

$$V^*(n) = \sum_{i=1}^{n} \sigma_i^2 \rho_i^2 - \lambda \sum_{i=1}^{n} \rho_i$$

Recursive form: $x_0(n+1) = x_0(n) + K_n[x_{n+1} - x_0(n)]$
Innovation, Kalman gain (matrix)


y and x0 vs time
y
2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
time
x10
3


y and x0 vs time
y
2.5
2.0
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
time
x10
4

ESTIMATION OF A CORRELATED HIGHER DIMENSIONAL VARIABLE

$$\mathbf{x}^i = \mathbf{x}_0 + \overrightarrow{\xi}^i$$

with $< \overrightarrow{\xi}^i >= 0 < \xi_k^i \xi_l^j >= \delta_{ij} C_{kl}^i$.

$$\chi^2 = \frac{1}{2} \sum_{i=1}^{n} (\overrightarrow{\xi}^i, \mathsf{D}^i \overrightarrow{\xi}^i)$$

$$= \frac{1}{2} \sum_{i=1}^{n} \left((\mathbf{x}^i - \mathbf{x}_0), \mathsf{D}^i (\mathbf{x}^i - \mathbf{x}_0) \right)$$

where $\mathsf{D}^i = (\mathsf{C}^i)^{-1}$.

$$\mathbf{x}_0 = \left(\sum_{i=1}^{n} \mathsf{D}^i \right)^{-1} \sum_{i=1}^{n} \mathsf{D}^i \mathbf{x}^i \qquad (1)$$

Note: this gives sample mean if all D_i are equal.

Also, if D_i are diagonal, this gives the weighted sample mean.

MINIMUM VARIANCE ALTERNATIVE – TRACE OF THE COVARIANCE MATRIX

$$\mathbf{x}_0(n) = \sum_{i=1}^{n} \mathsf{A}^i \mathbf{x}^i \qquad\qquad \sum_{i=1}^{n} \mathsf{A}^i = \mathsf{I}$$

$C(\overrightarrow{x}_0)_{kl} =< \delta x_{0,k} \delta x_{0,l} >$. Its *trace* is

$$T =< \sum_k \delta x_{0,k} \delta x_{0,k} >=< |\delta \mathbf{x}_0|^2 >$$

$$T = \sum_{ijkmn} A^i_{km} A^j_{kn} < \xi^i_m \xi^j_n >= \sum_{ikmn} A^i_{km} A^i_{kn} C^i_{mn}$$

$T = trace \sum_i \left(\mathsf{ACA}^T \right)^i$. Minimize T

$$T^* = \sum_{ikmn} A^i_{km} A^i_{kn} C^i_{mn} - \sum_{mn} \lambda_{mn} \left(\sum_i A^i_{mn} - \delta_{mn} \right)$$

Differentiating with respect to A^i_{ab} $\partial T^* / \partial A^i_{ab} = 0$

$$2A^i_{an} C^i_{nb} = \lambda_{ab} \qquad \mathsf{A} = \frac{1}{2} \mathsf{LD} \qquad \mathsf{L} = 2 \left(\sum_i D^i \right)^{-1}$$

$$\mathsf{A}^i = \left(\sum_i D^i \right)^{-1} \mathsf{D}^i$$

same as before

$$\mathbf{x}_0 = \left(\sum_{i=1}^n \mathsf{D}^i \right)^{-1} \sum_{i=1}^n \mathsf{D}^i \mathbf{x}^i \tag{2}$$

Random walk, estimating the current state – another alternate

Recall $y_k = x_n - \sum_{i=k}^{n-1} \xi_i = x_n - \zeta_k \ldots$

$$\chi^2 = \frac{1}{2} \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (x_n - y_k) D_{kl} (x_n - y_l) \qquad (3)$$

$$\mathsf{C} = \sigma_0^2 \begin{bmatrix} n-1 & n-2 & n-3 & \cdots & 1 \\ n-2 & n-2 & n-3 & \cdots & 1 \\ n-3 & n-3 & n-3 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \qquad (4)$$

Alternatively,

$$\chi^2 = \frac{1}{2\sigma_0^2} \sum_{k=1}^{n-1} \xi_i^2$$

$$= \frac{1}{2\sigma_0^2} \left[(y_2 - y_1)^2 + \ldots + (y_{n-1} - y_{n-2})^2 + (x_n - y_{n-1})^2 \right]$$

Obviously gives $x_n(n) = y_{n-1}$. $(\xi_0, \ldots, \xi_{n-1}) \rightarrow (\zeta_0, \ldots, \zeta_{n-1})$ – change of variable.

PROBABILISTIC (BAYESIAN) APPROACH – counting statistics

Bayes'

$$f(x_0|y_1) \propto f(y_1|x_0) \propto e^{-\frac{(y_1-x_0)^2}{2\sigma_1^2}}$$

$$f(x_0|y_1,y_2) \propto f(y_2|x_0,y_1)f(x_0|y_1)$$
$$f(y_2|x_0,y_1) = f(y_2|x_0)$$

$$f(x_0|y_1) \propto f(y_2|x_0)f(y_1|x_0)$$

Similarly $f(x_0|y_1,y_2,...,y_n) \propto f(y_n|x_0)\cdots f(y_2|x_0)f(y_1|x_0)$

$$\propto e^{-\frac{(y_n-x_0)^2}{2\sigma_n^2}} \cdots e^{-\frac{(y_2-x_0)^2}{2\sigma_2^2}} e^{-\frac{(y_1-x_0)^2}{2\sigma_1^2}}$$

Likelihood $\chi^2 = -\ln f \propto \sum_{k=1}^{n} \frac{(y_k-x_0)^2}{2\sigma_k^2}$ ... SAME

RANDOM WALK

$$f(x_0|y_1) \propto f(y_1|x_0) \propto e^{-\frac{(y_1-x_0)^2}{2\sigma_1^2}}$$

$$f(x_0|y_1, y_2) \propto f(y_2|x_0, y_1) f(x_0|y_1)$$

$$\propto f(y_2|y_1) f(x_0|y_1) \;\; (Markov) \;\; \propto f(y_2|y_1) f(y_1|x_0)$$

$$f(x_0|y_1, y_2, ..., y_n) \propto f(y_n|y_{n-1}) \cdots f(y_2|y_1) f(x_0|y_1)$$

$$f \propto e^{-\frac{(y_n-y_{n-1})^2}{2\sigma_{n-1}^2}} \cdots e^{-\frac{(y_2-y_1)^2}{2\sigma_1^2}} e^{-\frac{(y_1-x_0)^2}{2\sigma_0^2}}$$

$$\chi^2 = -\ln f \propto \sum_{k=1}^{n-1} \frac{(y_{k+1} - y_k)^2}{2\sigma_k^2} + \frac{(y_1 - x_0)^2}{2\sigma_0^2}$$

Random walk with measurement noise added – alternate approach

$$x_{k+1} = x_k + \xi_k,$$

$$y_k = x_k + \eta_k.$$

Then $y_{k+1} - y_k = \xi_k + \eta_{k+1} - \eta_k$ and

$$\chi^2 = \sum_{k=1}^{n-1} \sum_{l=1}^{n-1} (y_{k+1} - y_k) D_{kl} (y_{l+1} - y_l) \qquad y_n \to x_n(n)$$

with $C_{kl} =< (\xi_k + \eta_{k+1} - \eta_k)(\xi_l + \eta_{l+1} - \eta_l) >$ tridiagonal C =

$$\sigma_\eta^2 \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 0 & 0 & -1 & 2 \end{bmatrix} + \sigma_{\xi^2} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Minimize with respect to $x_n(n)$:

$$\left(x_n(n) - y_{n-1}\right) D_{n-1,n-1} + \sum_{k=1}^{n-2} D_{n-1,k} \left(y_{k+1} - y_k\right) = 0$$

Limit 1: no measurement noise $\sigma_\eta^2 = 0$ ... $\mathsf{C} = \sigma_\xi^2 \mathsf{I}$

or $\mathsf{D} = \sigma_\xi^{-2}\mathsf{I}$ ..... $\mathbf{x}_n(\mathbf{n}) = y_{n-1}$

Limit 2: no dynamical noise $\sigma_\eta^2 = 0$ ...

$$\mathsf{C} = \sigma_\eta^2 \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 0 & 0 & -1 & 2 \end{bmatrix}$$

gives sample mean

$$x_n(n) = \frac{1}{n-1} \sum_{k=1}^{n-1} y_k$$

Recall one dimensional system with measurement noise

Recall 1D system with measurement noise $< \eta_k \eta_l >= \delta_{kl}$

$$x_{k+1} = \gamma x_k,$$

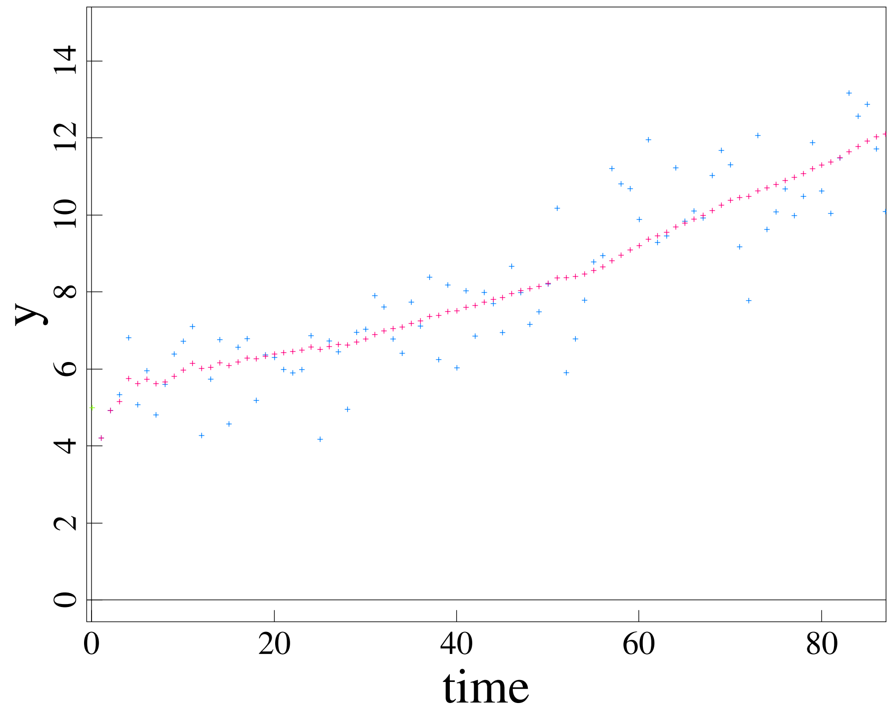
$$y_k = x_k + \eta_k.$$

$$x_0(n+1) = x_0(n) + \frac{\gamma^{n+1}}{\sum_{k=1}^{n} \gamma^{2k} + \gamma^{2n+2}} \left(y_{n+1} - \gamma^{n+1} x_0(n) \right).$$

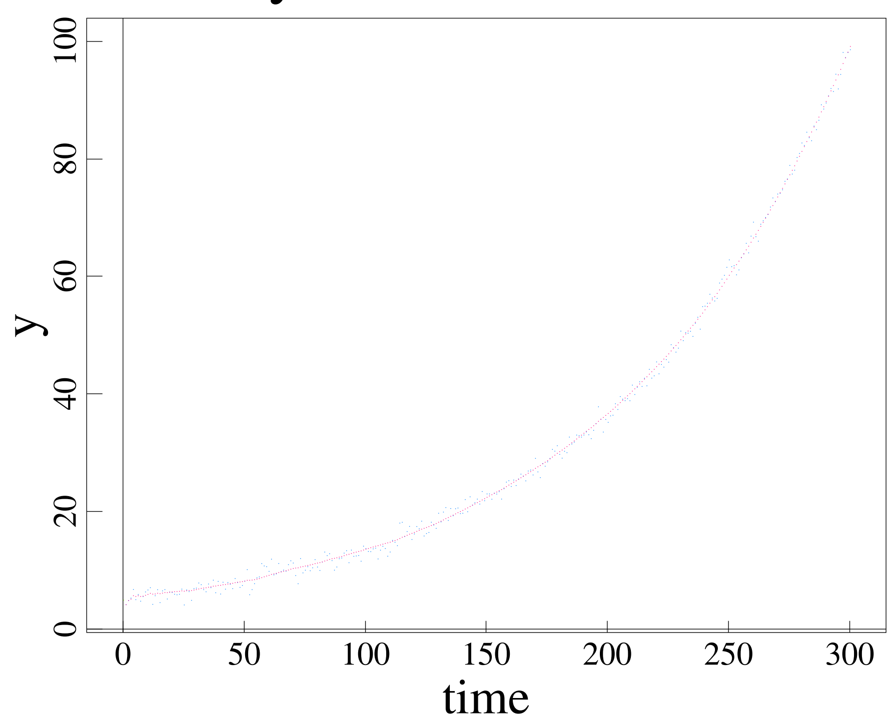
$$x_{n+1}(n+1) = \gamma x_n(n) + \frac{1}{\sum_{k=1}^{n} \gamma^{2k-2n-2} + 1} \left(y_{n+1} - \gamma x_n(n) \right).$$

Estimates $x_0(n)$ and $x_n(n) = \gamma^n x_0(n)$ ... $K_n^{x_n(n)} = \gamma^{n+1} K_n^{x_0(n)}$.

Also, $V(x_0) = \frac{1}{\sum_{k=1}^{n} \gamma^{2k}}$ $V(x_n) = \frac{\gamma^{2n}}{\sum_{k=1}^{n} \gamma^{2k}} = \gamma^{2n} V(x_0(n))$... Recursive y and x0 vs time



y and x0 vs time



Higher dimensional system with measurement noise - est. for $\mathbf{x}_0(n)$

$$\mathbf{x}_{k+1} = \mathsf{A}_k \mathbf{x}_k \tag{5}$$

with measurement $< \eta_k^i \eta_l^j >= \delta_{ij}\delta_{kl}$

$$\mathbf{y}_k = \mathsf{M}_k \mathbf{x}_k + \overrightarrow{\eta}_k. \tag{6}$$

$\mathbf{x}_k = \mathsf{U}_{k,0}\mathbf{x}_0 = \mathsf{A}_{k-1}\mathsf{A}_{k-2}...\mathsf{A}_0\mathbf{x}_0$

$$\chi^2 = \frac{1}{2}\sum_{k=1}^{n} \|\overrightarrow{\eta}_k\|^2 = \frac{1}{2}\sum_{k=1}^{n} \|\mathsf{M}_k \mathsf{U}_{k,0}\mathbf{x}_0 - \mathbf{y}_k\|^2, \tag{7}$$

$$\mathbf{x}_0(n) = \left[\sum_{k=1}^{n} \mathsf{N}_{k,0}^T \mathsf{N}_{k,0}\right]^{-1} \sum_{k=1}^{n} \mathsf{N}_{k,0}^T \mathbf{y}_k \qquad \mathsf{N}_{k,0} \equiv \mathsf{M}_k \mathsf{U}_{k,0}. \tag{8}$$

$$\mathbf{x}_0(n+1) = \mathbf{x}_0(n) + \mathsf{K}_n\left[\mathbf{y}_{n+1} - \mathsf{M}_{n+1}\mathsf{U}_{n+1,0}\mathbf{x}_0(n)\right] \quad (9)$$

$$\mathsf{P}_{n+1}^{-1} = \mathsf{P}_n^{-1} + \mathsf{N}_{n+1,0}^T\mathsf{N}_{n+1,0}$$

$$\mathsf{K}_n = \mathsf{P}_{n+1}\mathsf{N}_{n+1,0}^T \qquad \mathsf{C}(x_0(n)) = \mathsf{P}_n$$

$\mathbf{x}_0(n)$ propagated $n \to n+1$ by $\mathsf{U}_{n+1,0}$

Measurement applied $\mathsf{M}_{n+1}\mathsf{U}_{n+1,0}\mathbf{x}_0(n)$ is best guess for $\mathbf{y}_{n+1}$ before measurement

Higher dimensional system with measurement noise - est. for $\mathbf{x}_n(n)$

$$\chi^2 = \frac{1}{2}\sum_{k=1}^{n} \eta_k^2 = \frac{1}{2}\sum_{k=1}^{n} \|\mathsf{M}_k \mathsf{U}_{k,n}\mathbf{x}_n - \mathbf{y}_k\|^2$$

$$\mathbf{x}_n(n) = \left[\sum_{k=1}^{n} \mathsf{U}_{0,n}^T \mathsf{N}_{k,n}^T \mathsf{N}_{k,n} \mathsf{U}_{0,n}\right]^{-1} \sum_{k=1}^{n} \mathsf{U}_{0,n}^T \mathsf{N}_{k,n}^T \mathbf{y}_k, \tag{10}$$

or

$$\mathbf{x}_n(n) = \mathsf{U}_{n,0}\mathbf{x}_0(n) \quad \widetilde{\mathsf{K}}_n = \mathsf{U}_{n+1,0}\mathsf{K}_n$$

$$\mathbf{x}_{n+1}(n+1) = \mathsf{A}_n\mathbf{x}_n(n) + \widetilde{\mathsf{K}}_n\left[\mathbf{y}_{n+1} - \mathsf{M}_{n+1}\mathsf{A}_n\mathbf{x}_n(n)\right], \tag{11}$$

$\mathbf{x}_n(n)$ is advanced in time $\mathbf{x}_n(n) \to \mathsf{A}_n\mathbf{x}_n(n)$ and the measurement operation M_{n+1} is done. This is the best guess for $\mathbf{y}_{n+1}$ before $\mathbf{y}_{n+1}$ is measured

Continuous time advance, discrete time measurement formulation

$$\frac{d\mathbf{x}}{dt} = \mathsf{A}(t)\mathbf{x} + \overrightarrow{\xi}(t) \qquad <\xi_i\xi_j>= \mathsf{C}_0$$
$$\mathbf{y}_k = \mathsf{M}\mathbf{x}_k + \overrightarrow{\eta}_k \qquad <\eta_i\eta_j>= \mathsf{C}_1$$

1. Time advance of estimate and covariance between measurements

$$\frac{d\widehat{\mathbf{x}}}{dt} = \mathsf{A}(t)\widehat{\mathbf{x}} \quad \frac{dC}{dt} = \mathsf{AC} + \mathsf{CA}^T + \mathsf{C}_0$$

2. Adjust estimate and covariance at new measurement

$$\mathsf{K}_k = \mathsf{C}^{(-)}(t_k)\mathsf{M}^T\left[\mathsf{MC}^-(t_k)\mathsf{M}^T + \mathsf{C}_1\right]^{-1}$$
$$\mathsf{C}(t_k) = [\mathsf{I} - \mathsf{K}_k\mathsf{M}]\,\mathsf{C}^{(-)}(t_k)$$
$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^{(-)} + \mathsf{K}_k\left(\mathbf{y}_k - \mathsf{M}\widehat{\mathbf{x}}_k^{(-)}\right)$$

$\widehat{\mathbf{x}}_k^{(-)}$ is the best guess for $\mathbf{y}_k$ at t_k before its measurement; $\mathsf{C}^{(-)}(t_k)$ is the covariance matrix at t_k before measurement of $\mathbf{y}_k$.

Application to control theory – separation theorem

$$\mathbf{x}_{k+1} = \mathsf{A}_k \mathbf{x}_k + \overrightarrow{\xi}_k + \mathbf{u}_k \tag{12}$$

$$\mathbf{y}_k = \mathsf{M}_k \mathbf{x}_k + \overrightarrow{\eta}_k. \tag{13}$$

Continuum model...

$$\frac{d\mathbf{x}}{dt} = A(t)\mathbf{x} + \overrightarrow{\xi}(t) + \mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{x}(t) = \mathsf{M}(t)\mathbf{x}(t) + \overrightarrow{\eta}(t) \;\; \textit{special case}$$

Optimal control $\Rightarrow$minimizing for example

$$J = \int_0^T \{(\mathbf{x}(t), \mathsf{Q}(t)\mathbf{x}(t)) + (\mathbf{u}(t), \mathsf{R}(t)\mathbf{u}(t))\}\, dt$$

Minimizing J determines $\mathbf{u}[\mathbf{x}]$ optimally for $\overrightarrow{\xi}(t) = 0$. Degree of control vs. cost.

For $\overrightarrow{\xi}(t) \neq 0$ do the following:

- Find optimal control $\mathbf{u}[\mathbf{x}(t), t]$ for $\overrightarrow{\xi}(t) = 0$
- Use Kalman filter on $\mathbf{y}(t)$ to determine the optimal estimate $\widehat{\mathbf{x}}(t)$
- Add control $\mathbf{u}(\widehat{\mathbf{x}}(t), t)$ based on ***estimate*** to equation $d\mathbf{x}/dt = \ldots$

✘ Allows one to design controller and estimator independently

✘ More general form with measurement noise exists too

✘ A similar formulation exists for the discrete system

Extended Kalman Filter – for nonlinear systems

Most real problems (systems and measurements) are nonlinear

$$\frac{d\mathbf{x}}{dt} = \mathbf{a}(\mathbf{x}, t) + \overrightarrow{\xi}(t)$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \overrightarrow{\eta}(t)$$

- Advance the estimate between measurements by the non-linear dynamics

$$d\widehat{\mathbf{x}}/dt = \mathbf{a}(\widehat{\mathbf{x}}, t)$$

- Advance the covariance between measurements by

$$d\mathsf{C}/dt = \mathsf{A}(\widehat{\mathbf{x}}, t)\mathsf{C} + \mathsf{C}\mathsf{A}^T(\widehat{\mathbf{x}}, t) + \mathsf{C}_0$$

with $A_{ij} = \partial a_i / \partial x_j$ LINEARIZE with respect to $\mathbf{x}$

- Kalman gain

$$\mathsf{K}_k = \mathsf{C}^{(-)}(t_k)\mathsf{M}^T(\widehat{\mathbf{x}}_k^{(-)}) \times \left[\mathsf{M}(\widehat{\mathbf{x}}_k^{(-)})\mathsf{C}^{(-)}(t_k)\mathsf{M}^T(\widehat{\mathbf{x}}_k^{(-)}) + \mathsf{C}_1\right]^{-1}$$

  where $M_{ij} = \partial h_i/\partial x_j$.. LINEARIZE with respect to $\mathbf{x}$. Covariance similarly

- Update estimate after new data: $\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}_k^{(-)} + \mathsf{K}_k \left(\mathbf{y}_k - \mathbf{h}(\widehat{\mathbf{x}}_k^{(-)})\right)$

- Caveat: $d\widehat{\mathbf{x}}/dt = \mathbf{a}(\mathbf{x}, t) = \mathbf{a}(\widehat{\mathbf{x}}, t) + (\mathbf{x} - \widehat{\mathbf{x}}) \cdot \nabla \mathbf{a}(\widehat{\mathbf{x}}, t) + \ldots$

- Caveat: gaussian statistics remains gaussian only if C remains small –if linearizations hold over the range specified by C

- Caveat: what if the model [i.e. $\mathbf{a}(\mathbf{x}, t)$] is known poorly? Model errors

SUMMARY

- Least squares approach
- Recursive least squares. Kalman gain $\leftrightarrow$ covariance matrix; 'innovation'
- Minimum variance - minimum trace of the covariance matrix
- Estimating the initial state or the current state

  Only measurement noise – initial and current state estimates are related by the dynamics

  Only dynamical noise – initial and current state estimates are dominated by nearby data
- Bayesian approach and maximum likelihood $\longrightarrow$ least squares
- Higher dimension – principles the same (recursion for estimate and covariance matrix; relation with Kalman gain)
- Control theory and the separation theorem

- Extended Kalman Filter – advance estimate nonlinearly, covariance matrix by linearized system. Caveats:

  1) small covariance for linearization to be accurate ... otherwise not gaussian

  2) systematic errors – model errors